Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 151 (2019) 1213–1218

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Web Search and Data Mining (WSDM) April 29 - May 2, 2019, Leuven, Belgium

Privacy Preserving, Protection of Personal Data, and Big Data: a Review of the Colombia Case

Jesus Silva[a]*, Darwin Solano[b], Claudia Fernandez[c], Ligia Romero[d], Jesus Vargas Villa[e]

[a] *Universidad Peruana de Ciencias Aplicadas, Lima 07001, Peru.*
[b,c,d,e] *Universidad de la Costa (CUC), Barranquilla 080003, Colombia*

Abstract

Big Data promises great socially accepted and desirable benefits. However, in general terms, the datification of life has made people to lose some awareness of the risks involved in the massive analysis of data regarding their fundamental rights. This fact is used by the companies involved in the data value chain to maximize their benefits although this implies the proliferation of negative externalities assumed by the information holders. The Colombian State has made great efforts regarding the protection of data and privacy, as demonstrated by Law 1266 of 2008 and Law 1581 of 2012, nevertheless, a deep literary review leads to conclude the need to adapt to the international context.

Keywords: privacy preserving; personal protection; data and big data; sensitive information; Colombia.

1. Introduction

In the last decade, eight of the leading global economies, including the United States, the United Kingdom, Australia, South Korea, the European Union, France, Japan, and China have made progress in a national policy on data exploitation. This represents a very significant advance in Colombia, which is taking a step forward with this issue as reported by the Revista Dinero ("Big Data Public Policy", 2018) [1].

* Corresponding author. Tel.: +51920287620
E-mail address: jesussilvaUCP@gmail.com

10.1016/j.procs.2019.04.174

Colombia is the first country in Latin America to develop a public data exploitation policy through the National Council of Economic and Social Policy (CONPES - Consejo Nacional de Política Económica y Social), declared on April 17, 2018, thanks to the joint work of the National Planning Department, Ministry of Information, Technologies and Communications, and the Superintendence of Industry and Commerce. This policy is intended to invest around 16,728 million Colombian pesos to expand the use of data, so that they can be managed as assets to generate social and economic value (DNP, 2018) [2].

However, the expected benefits of the massive data analysis is not only referred to the public sphere, but extended to the private sector. The Economist newspaper affirms the existence of an "economy of data" where people exchange their data for free access to digital services, like the access to social media platforms or search engines. These data are collected, stored, and subjected to analytical processes that allow discovering valuable uses (Puldain, V. 2017)[3], (Supriyadi, D. 2017)[4].

At a global level, companies like Facebook, Google, Amazon, IBM are taking advantage of this technology. In Colombia there are companies like Proclive that, using Big Data are obtaining great benefits by optimizing business decisions, predicting the behavior of consumers to better satisfy their needs, improving marketing strategies, among other uses (Lane, J. et al; 2014)[5], (John, L et al; 2009)[6], (Fuster, G and Scherrer, A; 2015)[7], (Gellman, R. 2017)[8].

Nonetheless, Big Data involves the handling of personal information, as well as making (automated) decisions that affect the interests or rights of people. In Colombia, personal data are protected by the Political Constitution, Law 1266 of 2008, Law 1581 of 2012, constitutional jurisprudence, and other rules of a regulatory nature. All these rules are based on the fundamental right to habeas data (Fernández, J. 2017)[9], (Márquez, B. 2016)[10].

In this research, a literature review and analysis is carried out to obtain an understanding of the legal conditions in terms of protection of personal data that should be guaranteed in the massive data analysis (Big Data) to ensure the individual's computing self-determination.

2. Method

To address the problems proposed in the research, some methods of linguistic, logical, systematic, comparative, and empirical analysis are applied to materials such as doctrine, jurisprudence, and legal texts, as exposed by Daros, W (2017)[11], both for Colombia and other States, regarding the protection of personal data in the digital context.

The use of linguistic and logical analysis defines fundamental concepts such as habeas data, big data, computer power, principles of personal data management, as well as identifies the propositional relationships that exist between them, the logical problems, definition, coherence, and fullness of the legal order in relation to the data protection (Bayern, S; 2009)[12], (Daries, J; 2014)[13], (Garriga, D; 2016)[14], (Viloria, A and Viviana Robayo, P; 2016)[15], (Gaitán-Angulo M. et al; 2018)[16].

3. Results and Discussions

The national and international regulatory regime for the protection of personal data, as well as its scientific and philosophical basements have evolved according to the technological and economic conditions in which computer power is globally exercised. Since the middle of the 20th century, the appearance of the computer of centralized architecture (Albrecht, J; 2016)[17], which allowed the automated processing of personal information, permitted the appearance of the concepts of transparency, security, purpose, restricted circulation, independent authority, damage, and other legal principles within the legal systems of the European and Anglo-Saxon States (Canbay, Y and Sağıroğlu, S; 2017)[18], to be applied to the personal data processing.

The appearance of the personal computer and the internet generated a regulatory boom in the area of personal data at a global level (Barocas, S and Nissenbaum, H; 2014)[19], with the proliferation of international, internal, and community standards. Thus, the results demonstrated that law is intrinsically linked to the technological and economic context, and as a product of society and history, its influence embraces the conditions of technology, economy, and the society, creating relationships that respond to human anthropological needs in particular contexts in order to ensure the dignity of the person. This was precisely what happened to the normative regime of habeas data (Puccinelli, O; 2015)[20], (Leenes, R and Kosta, E; 2015)[21], (Leenes R; 2016)[22].

The Colombian State adopted the international and foreign policy intended to protect habeas data through the promulgation of the Political Constitution of 1991, the subsequent constitutional jurisprudence, Law 1266 of 2008, and Law 1581 of 2012. In this legal system, the centrality of the human person prevails since it exists for the purpose of guaranteeing individual control over personal information in order to limit the computing power held by large private companies or public institutions. This was the conclusion drawn from the analysis of current regulations on the subject (Fernández, J. 2017)[9], (Márquez, B. 2016)[10].

However, with the appearance, in the last decade, of the Big Data, serious difficulties are generated for the application and compliance with the legal principles of personal data protection since the underlying logic of this new technology is substantially different from the aims pursued through the Colombian and foreign regulatory regime for the habeas data protection (Mayer-Schönberger, V. and Cukier, K; 2013)[23]. According to the legal doctrine, the principles of data administration is obsolete in the new context of the digital economy, for this reason, the Colombian State seems not to have an adequate level of protection, at least from the instrumental point of view, as it was demonstrated in the second chapter (Manrique, G; 2015)[24], (Parraguez, K and Caldera, E. (2016)[25].

Facing this situation, international experience shows that several regulatory adjustments have been made at an external level (Gil, G; 2016)[26]. The most important of these improvements is the European Parliament's expedition within the European Union of the RGDP reinforcing the new legal provision with the principles of freedom, transparency, and demonstrated responsibility (Gantz J and Reinsel D; 2011)[27], (Crawford, K and Schultz, J; 2014)[28]. Likewise, it enshrined new rights in order to empower the personal data holders, trying to mitigate the asymmetric relationship they have with the large companies that are benefited from the exploitation of their personal information. The question that emerges regarding this aspect is whether the Colombian State should adopt to this new normative trend.

This issue is especially complex because, when analyzing the incentives that the Colombian regime for the protection of personal data generates in private agents (Floridi, L; 2002)[29], (El Emam, K and Álvarez, C; 2014)[30], it was concluded that this regime encourages or favors corporate strategies tending to evade compliance with the principles of personal data administration due to the high regulatory costs (Jaramillo. R; 2016)[31], and the logical contradiction between the Big Data business model and the scope of such principles.

In first place, given the current technological conditions, the scope of application of Law 1581 becomes too broad as all information generated by a person in the network would be identifiable or determinable, which extends the property or ownership of the data to all the existing information in the network, and reduces the useful effect of the normative regime. As an alternative to this increasingly broad regulatory environment, companies decide to apply severe anonymization techniques, in which they get trapped since this condition also means a significant loss of utility. This is called the utility paradox (Hueso, L; 2017)[32].

Only large companies or data barons, owners of mass databases, can easily overcome the paradox of utility since they can evade compliance of regulations through tracking techniques of the information owner in the network without needing to collect identification features or identifiable information of the subject, or failing that, they can anonymize without losing utility when applying re-identification techniques. In this sense, Law 1581 really privileges the dominant position of the data barons in the market, given that regulatory costs are essentially assumed by small,

medium, or new entrepreneurs who enter the Big Data economy, which represents a big problem to implement the recommendation nine (9) established by CONPES in the document 3920 of 2018 (Beavers, A; 2013)[33], (Bennett, C and Bayley, R; 2016)[34], (Zarsky, T; 2016)[35].

4. Conclusions

The Colombian State faces the challenge of protecting the information holders in a data market operated by corporate leviathans, owners of massive databases, which operate not exclusively within the scope of their jurisdiction but globally. This allows to easily evade the action of the personal data protection authorities, taking advantage, to a certain extent, of the lack of coordination of the different existing data protection regimes. Under these circumstances, public authorities fail in the attempt to guarantee the centrality of the person and their constitutional rights in the digital network.

The legislation on personal data, Law 1581 of 2011, instead of limiting the computing power, ends up generating the opposite effect in the context of Big Data. On the one hand, compliance with the treatment purpose principle prevents the emergence of secondary data markets that would be exploited by new competitors in the market, so its real effect is to empower or reinforce the position of domain held by the data barons, generating losses of competitiveness for entrepreneurs or small entrepreneurs who, by applying this principle in strict sense, could not use the collected personal information for secondary uses.

At the international level, the proposed regulatory solution to this problem is to open the way for the employ of personal information in secondary uses, which are not necessarily related to the main purposes of the treatment, provided that they are fair and are used in scientific or statistic research, without affecting (Zhang, X et al; 2016)[36], (Engan, M; 2017)[37], (Forgó, N et al; 2017)[38] in any way, the interests or rights of individuals, that is, that they are not a basis for making automated decisions with effects against human people. Thus, this solution intends to reduce the negative externalities or risks that arise from Big Data in the automated decision-making processes. In the same way, the emergence of secondary data markets for research purposes is promoted for the benefit of both public institutions and private companies.

It should be borne in mind that the task of adjusting the principles of data protection to the new technological context is not exclusive to the Colombian State, since the nature of the risks that derive from Big Data is global. The agents that benefit from the massive analysis operate without being subject to jurisdictional limits (Hox, J and Boeije, H; 2005)[39], (Gayo, M; 2017)[40], (Gellert, R; 2018)[41] therefore, it is necessary to make a call to the international community to take this issue seriously since it threatens the self-determination of individuals with the emergence of data bases of ruin (in the terms of Paul Ohm). Lastly, it is essential that the data protection authorities be strengthened at the global level so different data protection regimes can be coordinated, and effective solution models can be established to restore the centrality of the data owner.

References

[1] Política de Big Data. (2018). Revista Dinero. Recuperado de: https://www.dinero.com/pais/articulo/colombia-ya-tiene-politica-publica-de-big-data/257479.

[2] DNP. (2018). Documento CONPES 3920, Politica Nacional de Explotación de Datos (Big Data). Recuperado de: https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf

[3] Puldain Salvador, V. (2017). El futuro marco legal para la protección del acceso a los datos. Revista Ibero-Latinoamericana de Seguros, 26(47).

[4] Supriyadi, D. (2017). Personal and Non Personal Data, In the context of Big Data. Tesis, Tilburg. 123. Sriramoju, S. B. (2017). Introduction to big data: infrastructure and networking considerations. Recuperado de: http://one.com.vn/sites/default/files/file-attached/catalog/introduction_to_big_data__infrastructure_and_networking_considerations.pdf

[5] Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). Privacy, big data, and the public good: Frameworks for engagement. New York, NY: Cambridge University Press.

[6] John, L. K., Acquisti, A., & Loewenstein, G. (2009). The best of strangers: Context dependent willingness to divulge personal information. DOI: https://dx.doi.org/10.2139/ssrn.1430482

[7] Fuster, G. G., & Scherrer, A. (2015). Big Data and smart devices and their impact on privacy. Committee on Civil Liberties, Justice and Home Affairs (LIBE), Directorate-General for Internal Policies, European Parliament. Recuperado de: http://www. europarl. europa. eu/RegData/etudes/STUD/2015/536455/IPOL_STU (2015) 536455_EN. Pdf

[8] Gellman, R. (2017). Fair information practices: A basic history. Recuperado de: https://bobgellman.com/rg-docs/rg-FIPshistory.pdf

[9] Fernández, J. P. M. (2017). La protección de datos y los motores de búsqueda en Internet: Cuestiones actuales y perspectivas de futuro acerca del derecho al olvido/Data protection and Internet search engines: current issues and future perspectives about the right to be forgotten. Revista de Derecho Civil, 4(4), 181-209. Recuperado de: http://nreg.es/ojs/index.php/RDC/article/view/280/228.

[10] Márquez Buitrago, f. (2016). Aplicación de la ley estatutaria 1581 de 2012 a la red social facebook en colombia. Revista De Derecho Comunicaciones Y Nuevas Tecnologías, (15), 1-31. DOI:10.15425/redecom.15.2016.04

[11] Daros, W. R. (2017). ¿ Qué es un marco teórico?. Enfoques, 14(1 y 2), 73-112

[12] Bayern, S. J. (2009). Rational Ignorance, Rational Closed-Mindedness, and Modern Economic Formalism in Contract Law. California Law Review, 97(3), 943-973. DOI: http://dx.doi.org/https://doi.org/10.15779/Z38TD7R

[13] Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. Communications of the ACM, 57(9), 56-63. DOI: https://doi.org/10.1145/2643132

[14] Garriga Domínguez, A. (2016). Nuevos retos para la protección de datos personales. En la Era del Big Data y de la computación ubicua. Dykinson.

[15] Viloria, A., & Viviana Robayo, P. (2016). Virtual Network Level of Application Composed IP Networks Connected with Systems - (NETS Peer-to- Peer). Indian Journal Of Science And Technology, 9(46). doi:10.17485/ijst/2016/v9i46/107376

[16] Gaitán-Angulo M., Cubillos Díaz J., Viloria A., Lis-Gutiérrez JP., Rodríguez-Garnica P.A. (2018) Bibliometric Analysis of Social Innovation and Complexity (Databases Scopus and Dialnet 2007–2017). In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham

[17] Albrecht, J. P. (2016). The EU's New Data Protection Law–How A Directive Evolved Into A Regulation. Computer Law Review International, 17(2), 33-43. DOI: https://doi.org/10.9785/cri-2016-0202

[18] Canbay, Y., & Sağıroğlu, S. (2017). Big data anonymization with spark. In Computer Science and Engineering (UBMK), 2017 International Conference on (pp. 833-838). DOI: https://doi.org/10.1109/UBMK.2017.8093543

[19] Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. Privacy, big data, and the public good: Frameworks for engagement, 1, 44-75.

[20] Puccinelli, O. R. (2015). El hábeas data a veinte años de su incorporación en la Constitución argentina. Revista De Derecho Comunicaciones Y Nuevas Tecnologías, (13), 1-25. DOI:10.15425/redecom.13.2015.02

[21] Leenes, R., & Kosta, E. (2015). Taming the cookie monster with dutch law–a tale of regulatory failure. Computer Law & Security Review, 31(3), 317-335. DOI: https://doi.org/10.1016/j.clsr.2015.01.004

[22] Leenes R. (2016). Explaining the Unknown: Accountability and Transparency in Big Data Land. Data Science Seminar, Tilburg.

[23] Mayer-Schönberger, V., & Cukier, K. (2013). Big Data–A Revolution That Will Transform How We Live, Think and Work. Houghton Mifflin Harcourt.

[24] Manrique Gómez, V. (2015). El derecho al olvido: análisis comparativo de las fuentes internacionales con la regulación colombiana. Revista De Derecho Comunicaciones Y Nuevas Tecnologías, (14), 1-25. DOI:10.15425/redecom.14.2015.09

[25] Parraguez Kobek, L., & Caldera, E. (2016). Cyber Security and Habeas Data: The Latin American response to information security and data protection. Recuperado de: https://ssrn.com/abstract=2868039

[26] Gil Gonzales, Elena. (2016). Big Data, Privacidad y Protección de Datos. Accesit en el premio de investigación, Agencia Española de Protección de Datos.

[27] Gantz J, Reinsel D (2011). Extracting value from chaos. IDC iView, pp 1–12 Recuperado de: https://www.emcgrandprix.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

[28] Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev., 55, 93. DOI: https://ssrn.com/abstract=2325784

[29] Floridi, L. (2002). "Data", article for the International Encyclopedia of the Social Sciences, 2nd edition, editor in chief William A. Darity (Detroit: Macmillan)

[30] El Emam, K., & Álvarez, C. (2014). A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques. International Data Privacy Law, 5(1), 73-87. DOI: https://doi.org/10.1093/idpl/ipu033

[31] Jaramillo Romero, C. (2016). Derecho Fundamental al Hábeas Data:¿ Cómo se ha desarrollado y cuales han sido sus consecuencias en el ordenamiento jurídico colombiano?. Tesis de Grado. Recuperado de: https://repository.upb.edu.co/handle/20.500.11912/2877

[32] Hueso, L. C. (2017). Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales. Dilemata, (24), 131-150. Recuperado de: https://dialnet.unirioja.es/servlet/articulo?codigo=6066829

[33] Beavers, A. F. (2013). Floridi historizado: La cuestión del método, el estado de la profesión y la oportunidad de la filosofía de la información de Luciano Floridi. Escritos, 21(46), 39-68. Recuperado de: https://revistas.upb.edu.co/index.php/escritos/article/view/1782/1719

[34] Bennett, C. J., & Bayley, R. M. (2016). Privacy protection in the era of 'big data': regulatory challenges and social assessments. Exploring the Boundaries of Big Data, 205. Recuperado de: https://bartvandersloot.com/onewebmedia/Verkenning_32_Exploring_the_Boundaries_of_Big_Data.pdf#page=206

[35] Zarsky, T. Z. (2016). Incompatible: The GDPR in the Age of Big Data. Seton Hall L. Rev., 47, 995. Recuperado de: https://ssrn.com/abstract=3022646

[36] Zhang, X., Leckie, C., Dou, W., Chen, J., Kotagiri, R., & Salcic, Z. (2016). Scalable local-recoding anonymization using locality sensitive hashing for big data privacy preservation. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 1793-1802). ACM. DOI: https://doi.org/10.1145/2983323.2983841

[37] Engan, M. (2017). Big Data and GDPR. Tesis de maestría, University of Stavanger. Recuperado de: https://brage.bibsys.no/xmlui/handle/11250/2467489

[38] Forgó, N., Hänold, S., & Schütze, B. (2017). The Principle of Purpose Limitation and Big Data. In New Technology, Big Data and the Law (pp. 17-42). Springer, Singapore.

[39] Hox, J. J., & Boeije, H. R. (2005). Data collection, primary versus secondary. Encyclopedia of social measurement, 1, 593. Recuperado de: https://dspace.library.uu.nl/bitstream/handle/1874/23634/hox_05_data+collection,primary+versus+secondary.pdf?sequence=1

[40] Gayo, M. R. (2017). Big data: hacia la protección de datos personales basada en una transparencia y responsabilidad aumentadas. Revista De Derecho Comunicaciones Y Nuevas Tecnologías, (17), 1-24. DOI:10.15425/redecom.17.2017.09. 59.

[41] Gellert, R. (2018). Understanding the notion of risk in the General Data Protection Regulation. Computer Law & Security Review, 34(2), 279-288. DOI: https://doi.org/10.1016/j.clsr.2017.12.003 60.